Plasmodium falciparum accompanied the Human Expansion out of Africa.

Kazuyuki Tanabe[1]*, Toshihiro Mita[2], Thibaut Jombart[3], Anders Eriksson[4], Shun Horibe[5], Nirianne Palacpac[6], Lisa Ranford-Cartwright[7], Hiromi Sawai[1], Naoko Sakihama[1], Hiroshi Ohmae[8], Masatoshi Nakamura[9], Marcelo U. Ferreira[10], Ananias A. Escalante[11], Franck Prugnolle[12], Anders Björkman[13], Anna Färnert[14], Akira Kaneko[13, 14], Toshihiro Horii[6], Andrea Manica[4]*, Hirohisa Kishino[5], & Francois Balloux[3]*

[1]Laboratory of Malariology, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan; [2]Department of International Affairs and Tropical Medicine, Tokyo Women's Medical University, Tokyo 162-8666, Japan, [3]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Faculty of Medicine,Imperial College,London W2 1PG, United Kingdom;[4]Evolutionary Ecology Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK; [5]Graduate School of Agriculture and Life Sciences, University of Tokyo, Tokyo 113-8657, Japan; [6]Department of Molecular Protozoology, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan; [7]Division of Infection and Immunity, Faculty of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8TA, UK; [8]Department of Parasitology, National Institute of Infectious Diseases, Tokyo, Japan; [9]Department of Tropical Medicine and Parasitology, Dokkyo Medical University, Tochigi, Japan; [10]Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil; [11]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA; [12]IRD Montpellier, CNRS, IRD, Lab Genet & Evolut Malad Infect, UMR 2724, F-34394 Montpellier 5, France [13]Infectious Diseases Unit, Department of Medicine Solna, Karolinska Institutet, 17176 Stockholm, Sweden; [14]Global COE Program, Nagasaki University, Sakamoto, Nagasaki 852-8523, Japan;

* Corresponding authors

Running head: Population structure of *Plasmodium falciparum*

Summary

Plasmodium falciparum is distributed throughout the tropics and is responsible for an estimated 230 million cases of malaria every year with a further 1.4 billion people at risk of infection [1-3]. Little is known about the genetic makeup of *P. falciparum* populations, despite variation in genetic diversity being a key factor in morbidity, mortality and the success of malaria control initiatives. Here we analyze a worldwide sample of 519 *P. falciparum* isolates sequenced for two house-keeping genes (63 SNPs from around 5,000 nucleotides per isolate). We observe a strong negative correlation between within-population genetic diversity and geographic distance from Sub-Saharan Africa (R^2 = 0.95) over Africa, Asia and Oceania. In contrast, regional variation in transmission intensity and recent control initiatives seem to have had a negligible impact on the distribution of genetic diversity. The striking geographic patterns of isolation by distance observed in *P. falciparum* mirror the ones previously documented in humans [4-7] and point to a joint Sub-Saharan African origin between the parasite and its host. Age estimates for the expansion *P. falciparum* further support that anatomically modern humans were infected prior to their exit out of Africa and carried the parasite along during their colonization of the world, with the exception of the Americas which the parasite reached only more recently.

Results and Discussion

The genetic diversity of malaria parasites is central to their pathogenesis by facilitating immune evasion and drug resistance. As such, a better understanding of the worldwide distribution of the genetic diversity in *P. falciparum* is crucial for devising optimal drug- or vaccine- based malaria control strategies. Previous attempts at characterizing population structure in *P. falciparum* have pointed to considerable variation in genetic diversity and well differentiated populations over the parasite's vast distribution range [8-14]. Moreover, no consensus has been reached over worldwide trends in the distribution of genetic diversity, with either Africa or South America being identified as the region harboring the highest genetic diversity [8, 10, 13]. These inconsistencies between studies may be due to populations having been defined by pooling isolates sampled over heterogeneous geographic ranges. Alternatively, these discrepancies could be explained by the type of genetic markers that were deployed. Microsatellite markers may offer lower resolution for picking up patterns generated by relatively old demographic events due to their high mutation rate and high rate of back mutation (homoplasy). Conversely, Single Nucleotide Polymorphisms (SNPs) often suffer from biases induced by the selection of polymorphic markers on a small initial discovery panel. This phenomenon, referred to as ascertainment bias, leads to the genetic diversity of populations not well represented in the initial discovery panel being strongly underestimated [15-16].

The age of *P. falciparum* is also highly disputed [17-22], with previous estimates for the time to the most recent common ancestor (TMRCA) spanning more than an order of magnitude with values ranging from about 10,000 years [19, 22] to over 300,000 years [20]. With *P. falciparum* being an exclusively human parasite with no known animal reservoir, we hypothesized that if *P. falciparum* had been associated with humans for over 50,000 to 60,000 years (the estimated date for the out of Africa migration of anatomically modern humans), its current population structure could still carry a signal of human settlement history. Within-population genetic diversity of native human populations decreases smoothly with geographic distance measured through landmasses from a Sub-Saharan African origin [5], and genetic differentiation between populations also increases steadily with physical distance along landmasses [7, 23]. These smooth patterns in the distribution of human genetic diversity have been ascribed to sequential bottlenecks of small amplitude during the colonization of the world by our ancestors from an African cradle. In contrast, the parasite population structure may primarily depend on variation in epidemiological settings between populations. In particular, *P. falciparum* populations are characterized by high variability in variation in transmission intensity [1, 10, 18, 24-25], which could have affected local genetic diversity. Selective pressure imposed by anti-malaria interventions using drugs and insecticides might, likewise, have locally reduced genetic diversity.

To assess the relative importance of past human demography and recent epidemiological factors, we used a dataset of 519 *P. falciparum* isolates from nine

populations covering the entire distribution range of the parasite, with populations from Sub-Saharan Africa, South-East Asia, Oceania and South America. Care was taken to obtain a representative picture of the genetic makeup of *P. falciparum* populations by trying to minimize the geographic and temporal range of the isolates pooled into populations. We also chose to sequence two housekeeping genes, P-type Ca^{2+}-ATPase (*serca*) and adenylosuccinate lyase (*adsl*), in their entirety for all isolates to circumvent any possible ascertainment bias induced by marker selection.

We identified 49 and 14 SNPs in the *serca* (3630 bp) and *adsl* (1413 bp) genes, respectively (Supplementary Table 1). The ratio of synonymous (*dS*) over non-synonymous substitutions (*dN*) did not significantly deviate from neutral expectations overall and in any of the nine parasite populations considered individually (Supplementary Table 1). As there was no evidence for natural selection, we present all results on analyzes using both the synonymous and non-synonymous SNPs. However, using synonymous SNPs only did not affect any of the conclusions (results not shown). Comparison of average pair-wise nucleotide differences (θ_π) and standardized number of segregating sites (θ_S) indicates an excess of rare alleles in line with the high frequency of variants observed in a single population (private alleles) except in the two South American populations (Supplementary Table 1; Supplementary Figure 1).

Of the populations studied, those from Africa are by far the most genetically diverse, with the exception of Brazil, which is as diverse as Tanzania but only for the

θ_π estimate for the *serca* gene (Supplementary Table 1). However, we found no overall significant difference in within-population genetic diversity between the two genes. Thus, both genes were analyzed together in all subsequent analyses. The overall genetic differentiation is high with a global F_{ST} = 0.21. Between-population genetic distances (pair-wise F_{ST} estimates; Supplementary Figure 2) are highest for the two South American populations, which are strongly differentiated from all other populations, with the largest of all pair-wise differentiation observed between the two South-American populations (F_{ST} = 0.47; Supplementary Figure 2).

To test for an association between humans and *P. falciparum* predating the out of Africa exit, we computed the shortest distance through landmasses between each sampled population and a grid of hypothetical origins covering the entire world. We did not include the two South American populations in this analysis. The Americas were first colonized by humans some 15,000-20,000 years ago through the Bering Strait connecting Siberia to Alaska. The climate found today in the arctic is far too cold for the development of *P. falciparum*, which is unable to fulfill its life cycle at a temperature below 16-18°C [26-27]. As the temperature at the Bering Strait was considerably colder for the previous 90,000 years than it has been over the last 10,000 years [28], a joint colonization of *P. falciparum* together with humans into the Americas is highly implausible. Thus, the Americas have probably been colonized by this parasite far more recently, possibly through the slave trade [10, 17]. Interestingly, the resulting admixture could explain the relatively high level of parasite genetic diversity observed within American populations. We assumed that the cost of moving

along coastlines was half the friction of movement inland. We also allowed for a southern route of migration out of Africa [29], by creating land bridges on either side of the Arabic Peninsula at the Bab-el-Mandeb Strait and Hormuz Strait.

Using these geographic distances and within-population genetic diversities (θ_π), we searched for the hypothetical origin providing the best correlation between genetic diversity and the logarithm of geographic distance (Figure 1). We observed the highest correlation for a Central Sub-Saharan African origin (R^2 = 0.95; p < 0.0001; Figure 1). While this origin coincides with the one previously inferred for anatomically modern humans based on both genetic and morphological data [30], this result should be taken with some caution. The exact position of the inferred origin depends heavily on the genetic diversity of two African *P. falciparum* populations in the dataset. While the pattern is robust to the removal of either African populations (i.e. one at a time), ultimate confirmation of a coinciding geographic origin for both the parasite and its human host will require the inclusion of additional populations.

As previously reported in humans, there is also a strong isolation by distance (IBD) pattern with pair-wise genetic distances (pair-wise F_{ST}) being tightly correlated with geographic distance computed as above (Mantel r_M= 0.68; p < 0.001; Figure 2). The results are robust to the underlying assumptions in the analysis. We recover the same origin in Sub-Saharan Africa with a correlation of R^2 = 0.95 if we use θ_S instead of θ_π genetic diversity estimates. Similarly, assuming an equal cost between coastlines and inland or a three times lower cost for coastlines again points to a Sub-Saharan

African origin, with respective variance explained of 0.98 and 0.93 and IBDs of 0.46 and 0.66. Finally, forcing a northern route out of Africa through the Sinai into the Levant by closing the land bridges in and out of the Arabic Peninsula has a negligible effect on the best supported origin with a correlation between geography and genetic diversity of R^2 = 0.95 and an IBD of r_M= 0.68.

The smooth patterns in the apportionment of *P. falciparum* genetic diversity mirroring the ones previously described in native human populations are suggestive of an extensive association between the parasite and its host. However, this is not sufficient to conclude that humans were infected prior to their migration out of Africa and carried the parasite along during their colonization of the Old World. *P. falciparum* genetic diversity may be driven by a variable we did not consider but which is strongly correlated with distance from Africa. There are a variety of such candidate factors ranging from the availability of competent insect vectors to resistance alleles in the human host. As it is not feasible to consider sequentially a vast number of variables, we used instead the Entomological Inoculation Rate (EIR), an inclusive metric capturing current transmission intensity, which has been previously shown to correlate with *P. falciparum* genetic diversity as determined by microsatellite markers [10]. The intensity of malaria transmission varies greatly among geographic areas with the highest transmission levels recorded in Africa and Oceania (particularly on Papua New Guinea and the Solomon Islands) and much lower transmission intensity in Southeast Asia (Thailand and the Philippines) [25, 31]. EIR rates were obtained for all non-American populations from the literature (Supplementary Table 2). The correlation

between distance from Africa and genetic diversity remains highly significant after controlling for EIR (R^2 = 0.81, p = 0.0015), suggesting that sequential bottlenecks during the expansion out of Africa are the main determinant of the global patterns in within-population diversity of *P. falciparum*.

Alternatively, *P. falciparum* may have originated in Sub-Saharan Africa much more recently than anatomically modern humans but followed similar colonization routes to its human host during its expansion. To test for this possibility, we developed a new Approximate Bayesian Computation (ABC) framework of forward simulations of serial population expansions. This allowed us to estimate, among other parameters the timing of the initial expansion of *P. falciparum*, as well as colonization rates and subsequent migration between colonized demes.

To parameterize the simulations, we needed to obtain a substitution rate. To this end, we sequenced the two housekeeping genes in the chimpanzee malaria parasite *P. reichenowi* and estimated the substitution rate for the *serca* and *adsl* genes. There is no consensus on the exact timing of the split between *P. falciparum* and *P. reichenowi*. However, a range between around 2.5 Mya and 6 Mya is generally considered as credible [21, 32-34]. There have also been speculations that the split between the two parasites may have coincided with the advent of agriculture in West Asia, and could be as recent as 10,000 years ago [19, 22]. While the latter calibration date leads to an implausibly high substitution rate for eukaryotic coding genes of about 10^{-6}/site/year, we still decided to consider this hypothesis of a very recent host transfer.

We run four million simulations for each of the three substitution rates (based on the assumption of a 10,000 year, 2.5 Mya and 6 Mya split between to two malaria species). The 10,000 year split turned out to be highly implausible given the data, as we obtained not a single simulation with a reasonable fit to the observed pattern. Conversely, the model provided excellent fits between predicted and observed within population genetic diversity $\ln(\theta_\pi)$, with R^2~0.98 for the best combination of parameters for both scenarios with fast (split at 2.5 Mya) and slow mutation rate (split at 6 Mya). The model with a faster mutation rate pointed to the start of the spread out of Africa at around 40 Kya [90% CI: 19-77Kya] while the slower mutation rate gave 80-90 Kya [33-96 Kya] (Figure 3, Supplementary Table 3). These dates for the initial expansion of *P. falciparum* are compatible with the human out of Africa expansion some 50-60 Kya ago. Importantly the 90 % credibility intervals clearly exclude the advent of agriculture.

The demographic parameters were remarkably similar between the two scenarios (Supplementary Figure 3). The only exception was the size of the ancestral population (K_0) which was greater for the slower mutation rate. This is the result of the TMRCA being accounted for by a combination of the starting time of the expansion and the TMRCA in the founding population (determined by K_0). In general, both scenarios were characterized by a large ancestral population and relatively large carrying capacities for the demes during the expansion. Intense bottlenecks (Supplementary Figure 3) were needed to achieve the sharp decrease in within population genetic variability observed

in the data. Population growth rate (which defines the speed of expansion) was relatively fast. Conversely migration among demes was low, indicating that the pattern in genetic diversity was shaped primarily during the colonization of the tropics by *P. falciparum* rather than by subsequent exchanges between populations. Interestingly, the scenario we recover for the expansion of *P. falciparum* mirrors previous results for the human expansion out of Africa, with the same pattern of intense bottlenecks and low subsequent migration [35].

Our results establish that the genetic diversity of *P. falciparum* has been primarily shaped by human demography and does not provide evidence for a significant effect of contemporary nation-wide malaria interventions based on mass-drug administration and/or widespread spraying of DDT. For example, intensive malaria control programs in the Solomon Islands initiated in the 1970's and since discontinued, led to massive temporary reduction in parasite incidence [36]. Despite this, the genetic diversity found in the Solomon Islands does not fall below the curve for the genetic diversity as predicted by distance from Africa (Figure 1). The results also suggest that geographic variation in the distribution of insect vector species may have played only a minor role in shaping the population structure of *P. falciparum*, despite extensive variation in anthropophily and behavior between Anopheles species [25].

In this paper we have shown that the population genetic structure of *P. falciparum* outside of the Americas is primarily explained by geography, with 95% of the variance in within-population genetic diversity explained by physical distance from a Sub-

Saharan African origin alone. We further recovered strong patterns of isolation by distance and age estimates for the spread of *P. falciparum* coinciding with the colonization by anatomically modern humans of Africa, Asia and Oceania. Taken together, our results confirm that *P. falciparum* had already infected humans before the out-of-Africa migration and followed our ancestors in their expansion throughout the tropics, with the exception of the South America, which *P. falciparum* probably only reached much more recently through the slave trade. While an association between humans and *P. falciparum* predating the exit out of Africa has been suggested before [18, 20, 37], our results additionally demonstrate the dramatic impact of past human migrations on the current apportionment of genetic diversity in the parasite.

Experimental Procedures

DNA sample collection

P. falciparum isolates were collected from nine countries: Tanzania, Ghana, Thailand, the Philippines, Papua New Guinea (PNG), the Republic of Vanuatu, the Solomon Islands, Brazil, and Venezuela. Details of the samples from Tanzania, Thailand, Philippines, Solomon Islands, Vanuatu, and Brazil used in this study have been described previously [31, 38-44]. Briefly, in Tanzania, blood samples were collected from infected individuals in the Rufiji River Delta in eastern coastal Tanzania in 1993, 1998, and 2003 [40]. In Thailand, blood samples were collected from malaria patients attending a malaria clinic in Mae Sot near the northwestern Thailand-Myanmar border in 1995 [42]. In the Philippines, samples were collected from malaria patients attending hospitals in Palawan Island in 1997 [31]. In the Solomon Islands, samples were collected from infected individuals in northeastern Guadalcanal Island in 1995 – 1996 [39]. In Vanuatu, samples were collected from four islands, Malakula, Gaua, Esprit Santo and Pentecost between 1996 and 1998 [38, 42]. In Brazil, isolates were collected from malaria patients in five states: Acre, Rondonia, Mato Grosso, Para and Amapa between 1985 and 1999 [43]. Additional samples were collected from Acre in 2004 – 2005 [44]. Venezuelan isolates were collected from Upper Orinoco, Amazonas State, Venezuela in 1997 (Escalante, unpublished). In all cases, ethical clearance for sampling was obtained from relevant ethical committees. Isolates from Ghana and PNG were collected specifically for this study. In Ghana, 182 *P. falciparum*-infected

blood samples were collected during malaria surveys from 0-15 year old children in three villages (Okyereko, Mpota and Apam) near Winneba, a western coastal region in November 2004. The study was approved by the Ministry of Health/Ghana Health Service. In PNG, 195 malaria infected blood samples were collected during surveys in five villages in Wewak, East Sepik Province in northeast coast: Kiniambu in August 2001, Jawia and Witupe in September 2001, and Boiken and Wingei in Feburuary 2002. The study was approved by the National Department of Health Medical Research Advisory Committee of PNG and the Tokyo Women's Medical University Ethical Committee. Informed consent was obtained from the patients or their parents. In both Ghana and PNG, finger-prick blood was collected on Whatman® 31ETCHR filter paper. Parasite genomic DNA was extracted from filter blots using the EZ1 DNA Investigator kit on the EZ1 BioRobot™ (Qiagen, Germany).

DNA sequencing

Isolates infected with mixed *msp1* haplotypes, as determined by PCR-based haplotyping [31] were excluded from further analysis and only those with a single *msp1* haplotype infection (n = 519), were used for sequencing. Full-length sequences were obtained for two housekeeping genes, Ca^{2+}-transporting ATPase gene (*serca*) and adenylosuccinate lyase gene (*adsl*). The Ca^{2+}-transporting ATPase of *P. falciparum* has recently been suggested to be a potential target of artemisiins [45], anti-malarial drugs currently widely used for treatment in many endemic countries. However, all isolates in this study were collected before the adoption of this drug. Genomic DNA was

subjected to two independent PCR amplifications, the products of which were directly sequenced in both directions. Procedures and conditions used for PCR amplification and sequencing of *serca* have been described elsewhere [46]. For amplification of *adsl*, the first PCR was run using primers ASL-F3 (5'-TATAACTCCCCAAAACAAAACCACTAAAATGT) and ASL-R4 (5'-AAAGGCGTACATGTTATAAGGTCCT), followed by nested PCR using primers ASL-F2 (5'-ATTTATATATATTCCTTATTATATAGTCA) and ASL-R3 (5'-TGGGAGTGCCCAACTTGCAGTGTCT). Full-length *adsl* sequence was also obtained from *P. reichenowi*, a chimpanzee malaria parasite, closely related to *P. falciparum*. Whole genome-amplified *P. reichenowi* genomic DNA [44] was subjected to PCR using three primer sets targeting the 5'-, central- and 3'-regions of the gene: ASL-F3 and ASL-R6 (5'-CAATTATATAAGCATAAACCATATGCT), ASL-F6 (5'-AAATTGGAAGTAGTACCATGCCACA) and ASL-R4, and ASL-F5-2 (5'-GACAAATCATGATGTTAAGGCGGTTGA) and ASL-R5-2 (5'-CTTATTGGTAATTTGGAACTAAATAACTTGA). PCR conditions were identical for *serca*. Whenever there was an inconsistency between two sequences after independent amplifications, a third round of PCR/sequencing was performed. Only isolates with a single genotype infection, as judged by the lack of overlapping peaks on electropherograms, were used for analysis. The sequences reported in this study have been deposited in the DDBJ/EMBL/GenBank database (accession nos. AB501575-AB502442, AB AB519183, AB520081-AB520237).

Statistical genetics analyses

Two DNA sequence regions of Asn-codon tandem repeats in *serca* were excluded from the analyses. Nucleotide diversity was estimated by θ_π, the average pair-wise nucleotide distance, and θ_S, the standardized number of polymorphic sites per site (Watterson's estimator) using DnaSP version 4.10 [47]. A Z test was applied to test for a difference of in θ_S and θ_π between populations. The mean number of synonymous substitutions (*dS*) and non-synonymous substitutions (*dN*) were estimated by the Nei and Gojobori method [48] with a Jukes and Cantor correction as implemented in the MEGA software v. 3.1 [49]. Standard error was determined by 1,000 bootstrap replications, and *dN* and *dS* were compared with a Z test using MEGA. The inter-population variance in allele frequencies, F_{ST}, was calculated using Weir and Cockerham's θ estimator [50] with Arlequin v. 3.1 [51].

Spatially explicit analyses

All geographic distances were computed using graph theory [5, 23] as shortest distances along landmasses within a spherical referential of 40,962 vertices. We assumed that the friction cost along coastlines was half that of moving inland. Land bridges were created between the Malay Peninsula and Australia, connecting the major Indonesian islands. We further assumed two land bridges on either side of the Arabic peninsula at the Bab-el-Mandeb Strait and Hormuz Strait to allow for a Southern route of colonization out of Africa [29]. The best supported origin was

inferred by searching for the shortest routes to all analyzed populations from 312 hypothetical origins on land arranged on a regular grid. We controlled for Entomological Inoculation Rates (EIRs) using partial correlation. We tested whether within population diversity was still significantly correlated to geographic distance from Africa, once EIRs were accounted for first. For the isolation by distance analysis, a matrix of pair-wise physical distances was computed for all populations using the distance through landmasses described above.

Approximate Bayesian Computation

We modelled the expansion of malaria out of Africa by considering a one-dimensional stepping stone, an approach successfully used to investigate the out of Africa expansion of anatomically modern humans as well as *Helicobacter pylori* [6-7, 35, 52]. Although the real expansion would have followed a two-dimensional spread, numerical studies suggest that a one-dimensional framework is a reasonable approximation as long as the distance between samples is large. We used 300 demes of equal sizes, to represent the coastal route from the origin in Africa to Oceania (c.f. Figure 1). In order to avoid potential boundary effects at the origin, we added 50 demes to the African end (making the origin deme 51 in the chain).

Before the spread of *P. falciparum*, we imagine a well mixed population of size K_0 that represents the early presence of *P. falciparum* in Africa. At the onset of the spread, a seed population of size c_0K_0 was placed at the origin. This population increased linearly

with rate *r* until it reached size *K*, the carrying capacity for all demes in the simulation (in other words, *K* is the effective population size of a deme at carrying capacity). At each time step, demes that had reached their maximum size were allowed to send out colonists to adjacent empty demes and migrants to adjacent demes that had already been colonized. The number of colonists was given by the colonization rate *c* multiplied by the deme size *K*, and the number of migrants by the migration rate *m* multiplied by *K* (migrants were shared equally between the two adjacent neighbors). Once a deme was colonized by some individuals, its population increased linearly with growth rate *r* until it reached its maximum size *K*. *P. falciparum* undergoes about six generations a year [10]. After testing that generation times (scaled for mutations) up to sixty times slower (1/10 years) had no qualitative impact on the simulation results, we settled for a computationally reasonable compromise of one generation per year.

From the demography described above, we generated gene genealogies for the two unlinked genes (*serca* 3630 bp and *adsl* 1413 bp) according to the Wright-Fisher model: individuals were assumed to be randomly mating within each deme, and generations were non-overlapping. Assuming a Poisson process, we then simulated mutations on the gene genealogies; three mutations rates, 9.18×10^{-7}, 3.67×10^{-9} and 1.53×10^{-9}/site/year, were estimated from the data assuming a divergence between *P. falciparum* and *P. reichenowi* at 10,000, 2.5 million or 6 million years ago [19, 21-22, 32-34]. For each simulated gene genealogy, we then computed for each deme θ_π, the average number of pairwise differences per site between sequences within a deme. Rather than using the raw values of θ_π, which tend to have a highly skewed

distribution, we took natural logarithms of this variable, giving us an observable with an approximately symmetrical distribution.

We estimated the best parameter values that describe the spread of *P. falciparum* using Approximate Bayesian Computation (ABC) including weighted local regression [53], using within population $\ln(\theta_\pi)$ as our summary statistics. The stepping-stone parameter values for each simulation were sampled from uniform prior distributions of the log values of the following ranges: K $10–10^6$, K_0 $10–10^6$, m $10^{-6}–0.5$, c $10^{-6}–0.5$, c_0 $10^{-6}–1$, r $10^{-3}–1$. The start of the simulation t had a uniform distribution from 1,000 to 100,000 years ago. We also enforced two constraints: $c\,K \geq 1$ and $c_0\,K_0 \geq 1$. We run four million simulations for each of the three mutation rates, with an acceptance criterion of $R^2>0.7$ between observed and predicted $\ln(\theta_\pi)$. For the fastest mutation rate (referring to a split between the *Plasmodium* species 10,000 years ago), we were unable to obtain any simulation that fitted the data ($R^2>0.7$). For this reason, no results are shown, as this mutation rate is implausible given the data. For the other two mutation rates, we accepted 6,352 and 8,943 out of four million simulations for the split at 2.5 and 6 Mya, respectively.

Acknowledgements

This study is dedicated to the late Professor David Walliker. We thank all those who participated in the epidemiological studies for their kind cooperation, particularly I. Rooth, M. Dzodzomenyo, H. Eto, T. Tsukahara, F. Hombhanje, H. Osawa, I. Hwaihwanje, K. Hirayama, K. Na-Bangchang, A. Palanca Jr., R. Espina, and B, Bakote'e. We also want to acknowledge three anonymous reviewers for insightful suggestions. This work was supported by MEXT (18073013), JSPS (18GS03140013, 20390120), MHLW (H21-Shinkou-ippan) and the Takeda Science Foundation, Japan. AE and AM were funded by the Leverhulme Trust. FB acknowledges financial support from the Biotechnology and Biological Sciences Research Council and the Medical Research Council.

References

1. Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R.F., Brooker, S., et al. (2009). A World Malaria Map: Plasmodium falciparum Endemicity in 2007. PLos Med. *6*, 17.
2. Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y., and Hay, S.I. (2005). The global distribution of clinical episodes of Plasmodium falciparum malaria. Nature *434*, 214-217.
3. World_Health_Organisation (2008). World Malaria Report. Volume WHO/HTM/GMP/2008.1.
4. Handley, L.J.L., Manica, A., Goudet, J., and Balloux, F. (2007). Going the distance: human population genetics in a clinal world. Trends in Genetics *23*, 432-439.
5. Prugnolle, F., Manica, A., and Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. Curr Biol *15*, R159-160.
6. Liu, H., Prugnolle, F., Manica, A., and Balloux, F. (2006). A geographically explicit genetic model of worldwide human-settlement history. Am J Hum Genet *79*, 230-237.
7. Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. U. S. A. *102*, 15942-15947.
8. Mu, J.B., Awadalla, P., Duan, J.H., McGee, K.M., Joy, D.A., McVean, G.A.T., and Su, X.Z. (2005). Recombination hotspots and population structure in Plasmodium falciparum. PLoS Biology *3*, 1734-1741.
9. Mu, J.B., Awadalla, P., Duan, J.H., McGee, K.M., Keebler, J., Seydel, K., McVean, G.A.T., and Su, X.Z. (2007). Genome-wide variation and identification of vaccine targets in the Plasmodium falciparum genome. Nature Genetics *39*, 126-130.
10. Anderson, T.J.C., Haubold, B., Williams, J.T., Estrado-Franco, J.G., Richardson, L., Mollinedo, R., Bockarie, M., Mokili, J., Mharakurwa, S., French, N., et al. (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol. Biol. Evol. *17*, 1467-1482.
11. Jeffares, D.C., Pain, A., Berry, A., Cox, A.V., Stalker, J., Ingle, C.E., Thomas, A., Quail, M.A., Siebenthall, K., Uhlemann, A.C., et al. (2007). Genome variation and evolution of the malaria parasite Plasmodium falciparum. Nature Genetics *39*, 120-125.
12. Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner, D.A., Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O., et al. (2007). A genome-wide map of diversity in Plasmodium falciparum. Nature Genetics *39*, 113-119.
13. Neafsey, D.E., Schaffner, S.F., Volkman, S.K., Park, D., Montgomery, P., Milner, D.A., Lukens, A., Rosen, D., Daniels, R., Houde, N., et al. (2008). Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. Genome Biol. *9*, 16.

14. Volkman, S.K., Barry, A.E., Lyons, E.J., Nielsen, K.M., Thomas, S.M., Choi, M., Thakore, S.S., Day, K.P., Wirth, D.F., and Hartl, D.L. (2001). Recent origin of Plasmodium falciparum from a single progenitor. Science *293*, 482-484.
15. Rogers, A.R., and Jorde, L.B. (1996). Ascertainment bias in estimates of average heterozygosity. Am J Hum Genet *58*, 1033-1041.
16. Romero, I.G., Manica, A., Goudet, J., Handley, L.L., and Balloux, F. (2009). How accurate is the current picture of human genetic variation[quest]. Heredity *102*, 120-126.
17. Hartl, D.L. (2004). The origin of malaria: mixed messages from genetic diversity. Nat Rev Micro *2*, 15-22.
18. Joy, D.A., Feng, X.R., Mu, J.B., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E., et al. (2003). Early origin and recent expansion of Plasmodium falciparum. Science *300*, 318-321.
19. Rich, S.M., Light, M.C., Hudson, R.R., and Ayala, F.J. (1998). Malaria's eve: Evidence of a recent population bottleneck throughout the world populations of Plasmodium falciparum. Proceedings of the National Academy of Sciences of the United States of America *95*, 4425-4430.
20. Hughes, A.L., and Verra, F. (2001). Very large long-term effective population size in the virulent human malaria parasite Plasmodium falciparum. Proc. R. Soc. Lond. Ser. B-Biol. Sci. *268*, 1855-1860.
21. Escalante, A.A., and Ayala, F.J. (1994). Phylogeny of the malaria genus *Plasmodium*, derived from rRNA gene sequences. Proc. Natl. Acad. Sci. USA *91*, 11373-11377.
22. Rich, S.M., Leendertz, F.H., Xu, G., LeBreton, M., Djoko, C.F., Aminake, M.N., Takang, E.E., Diffo, J.L.D., Pike, B.L., Rosenthal, B.M., et al. (2009). The origin of malignant malaria. Proceedings of the National Academy of Sciences of the United States of America *106*, 14902-14907.
23. Manica, A., Prugnolle, F., and Balloux, F. (2005). Geography is a better determinant of human genetic differentiation than ethnicity. Hum Genet *118*, 366-371.
24. Kelly-Hope, L.A., and McKenzie, F.E. (2009). The multiplicity of malaria transmission: a review of entomological inoculation rate measurements and methods across sub-Saharan Africa. Malar. J. *8*, 16.
25. Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S.E., and Sachs, J. (2004). A global index representing the stability of malaria transmission. Am. J. Trop. Med. Hyg. *70*, 486-498.
26. Molineaux, L. (1988). The epidemiology of human malaria as an explanation of its distribution, including some implications for its control. In Malaria: principles and practice of malariology, W.H. Wernsdorfer and I. McGregor, eds. (Edinburgh: Churchill Livingstone), pp. 913-998.
27. Onori, E., and Grab, B. (1980). Indicators for the forecasting of malaria epidemics. Bull. World Health Organ. *58*, 91-98.
28. De Boer, A.M., and Nof, D. (2004). The Bering Strait's grip on the northern hemisphere climate. Deep-Sea Res. Part I-Oceanogr. Res. Pap. *51*, 1347-1366.
29. Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science *308*, 1034-1036.

30. Manica, A., Amos, W., Balloux, F., and Hanihara, T. (2007). The effect of ancient population bottlenecks on human phenotypic variation. Nature *448*, 346-U346.
31. Sakihama, N., Nakamura, M., Palanca, A.A., Argubano, R.A., Realon, E.P., Larracas, A.L., Espina, R.L., and Tanabe, K. (2007). Allelic diversity in the merozoite surface protein 1 gene of Plasmodium falciparum on Palawan Island, the Philippines. Parasitol. Int. *56*, 185-194.
32. Ollomo, B., Durand, P., Prugnolle, F., Douzery, E., Arnathau, C., Nkoghe, D., Leroy, E., and Renaud, F. (2009). A New Malaria Agent in African Hominids. PLoS Pathog *5*, e1000446.
33. Hayakawa, T., Culleton, R., Otani, H., Horii, T., and Tanabe, K. (2008). Big bang in the evolution of extant malaria parasites. Molecular Biology and Evolution *25*, 2233-2239.
34. Krief, S., Escalante, A.A., Pacheco, M.A., Mugisha, L., André, C., Halbwax, M., Fischer, A., Krief, J.-M., Kasenene, J.M., Crandfield, M., et al. (2010). On the Diversity of Malaria Parasites in African Apes and the Origin of *Plasmodium falciparum* from Bonobos. PLoS Pathog *6*, e1000765.
35. Deshpande, O., Batzoglou, S., Feldman, M.W., and Cavalli-Sforza, L.L. (2009). A serial founder effect model for human settlement out of Africa. Proceedings of the Royal Society B-Biological Sciences *276*, 291-300.
36. Over, M., Bakote'e, B., Velayudhan, R., Wilikai, P., and Graves, P.M. (2004). Impregnated nets or DDT residual spraying? Field effectiveness of malaria prevention techniques in Solomon Islands, 1993-1999. Am. J. Trop. Med. Hyg. *71*, 214-223.
37. Escalante, A., Barrio, E., and Ayala, F. (1995). Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. Mol Biol Evol *12*, 616-626.
38. Tanabe, K., Sakihama, N., and Kaneko, A. (2004). Stable SNPs in malaria antigen genes in isolated populations. Science *303*, 493-493.
39. Sakihama, N., Ohmae, H., Bakote'e, B., Kawabata, M., Hirayama, K., and Tanabe, K. (2006). Limited allelic diversity of Plasmodium falciparum merozoite surface protein 1 gene from populations in the Solomon Islands. Am. J. Trop. Med. Hyg. *74*, 31-40.
40. Tanabe, K., Sakihama, N., Rooth, I., Bjorkman, A., and Farnert, A. (2007). High frequency of recombination-driven allelic diversity and temporal variation of Plasmodium falciparum MSP1 in Tanzania. Am. J. Trop. Med. Hyg. *76*, 1037-1045.
41. Sakihama, N., Kaneko, A., Hattori, T., and Tanabe, K. (2001). Limited recombination events in merozoite surface protein-1 alleles of Plasmodium falciparum on islands. Gene *279*, 41-48.
42. Sakihama, N., Kimura, M., Hirayama, K., Kanda, T., Na-Bangchang, K., Jongwutiwes, S., Conway, D., and Tanabe, K. (1999). Allelic recombination and linkage disequilibrium within Msp-1 of Plasmodium falciparum, the malignant human malaria parasite. Gene *230*, 47-54.
43. Silva, N.S., Silveira, L.A., Machado, R.L.D., Povoa, M.M., and Ferreira, M.U. (2000). Temporal and spatial distribution of the variants of merozoite surface protein-1 (MSP-1) in Plasmodium falciparum populations in Brazil. Ann. Trop. Med. Parasitol. *94*, 675-688.

44. Ferreira, M.U., Karunaweera, N.D., da Silva-Nunes, M., da Silva, N.S., Wirth, D.F., and Hartl, D.L. (2007). Population structure and transmission dynamics of plasmodium vivax in rural amazonia. J. Infect. Dis. *195*, 1218-1226.
45. Eckstein-Ludwig, U., Webb, R.J., van Goethem, I.D.A., East, J.M., Lee, A.G., Kimura, M., O'Neill, P.M., Bray, P.G., Ward, S.A., and Krishna, S. (2003). Artemisinins target the SERCA of Plasmodium falciparum. Nature *424*, 957-961.
46. Tanabe, K., Sakihama, N., Hattori, T., Ranford-Cartwright, L., Goldman, I., Escalante, A.A., and Lal, A.A. (2004). Genetic distance in housekeeping genes between Plasmodium falciparum and Plasmodium reichenowi and within P-falciparum. Journal of Molecular Evolution *59*, 687-694.
47. Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics *19*, 2496-2497.
48. Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution *3*, 418-426.
49. Kumar, S., Tamura, K., and Nei, M. (2004). Mega3: Integrated software for Molecular Evolutionary Genetics analysis and sequence alignment. Briefings in Bioinformatics *5*, 159-163.
50. Weir, B., and Cockerham, C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358-1370.
51. Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online *1*, 47-50.
52. Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S.W., et al. (2007). An African origin for the intimate association between humans and Helicobacter pylori. Nature *445*, 915-918.
53. Beaumont, M., Zhang, W., and Balding, D. (2002). Approximate Bayesian computation in population genetics. Genetics *162*, 2025-2035.

FIGURE 1. **Map of most likely origin for *P. falciparum*.** Grey circles represent the geographic locations of the populations analyzed, with the surface of the circles proportional to within-population genetic diversity (θ_π). The background color represents the strength of the correlation between geographic distance from various origins and genetic diversity, with light shades representing more likely origins. The best supported origin is indicated by a blue dot. We also represented the colonization routes from this origin through landmasses to all populations analyzed (see methods). The inset represents the correlation between geographic distance (measured as travel cost over friction routes) and genetic diversity for the best supported origin.

FIGURE 2. **Plot of isolation by distance.** Logarithm of pair-wise geographic distances along landmasses (measured as travel costs over the friction routes) along and pair-wise genetic distances (F_{ST}) between populations.

FIGURE 3. **Posterior distributions of the date of expansion of *P. falciparum*.** These curves represent the posterior distributions of the date of expansion of *P. falciparum* obtained by the ABC analysis. The two curves have been generated under the assumption of a split between the two Plasmodium species (*P. falciparum and P. reichenowi*) at 2.5 (blue) and 6 (red) Mya.

Figure 1

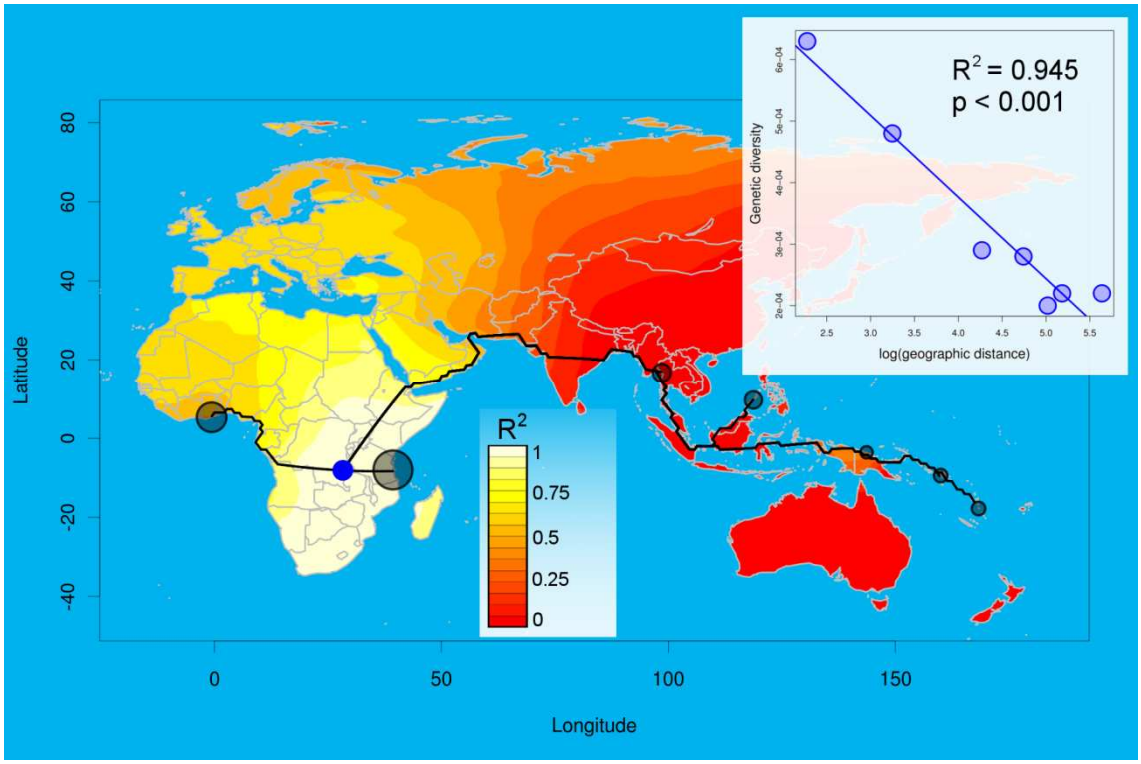

R² = 0.945
p < 0.001
Genetic diversity
log(geographic distance)
R²
1
0.75
0.5
0.25
0
Latitude
Longitude
80
60
40
20
0
-20
-40
0
50
100
150

Figure 2

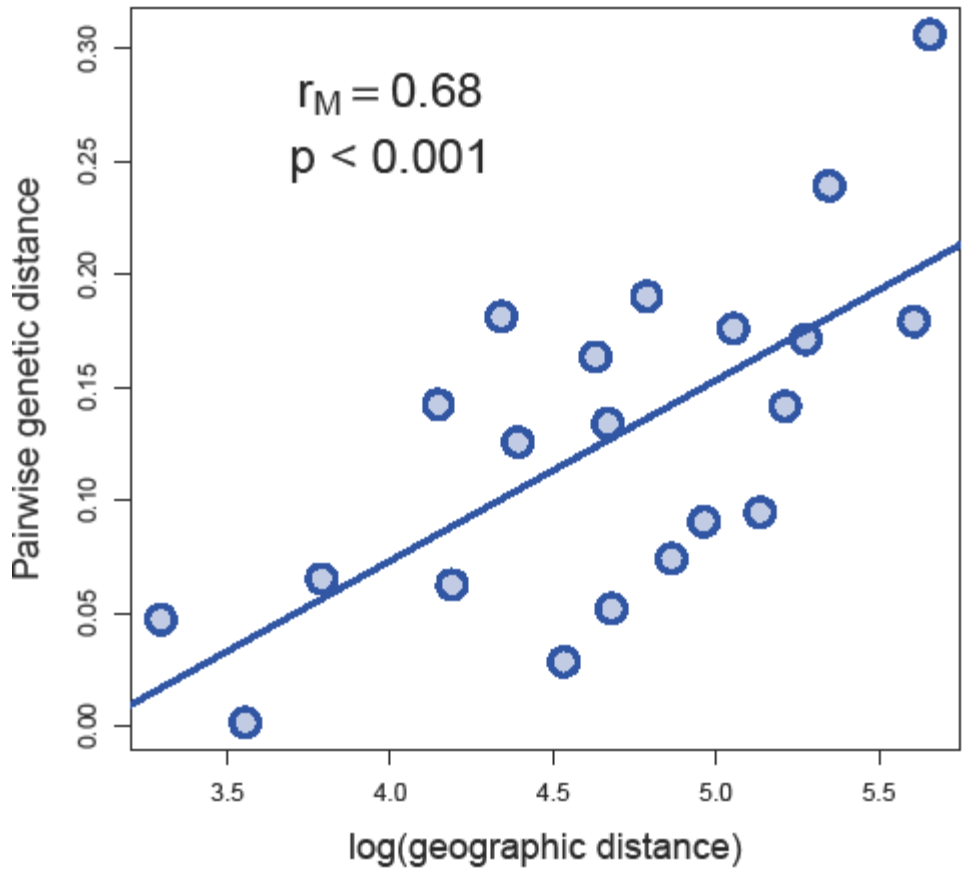
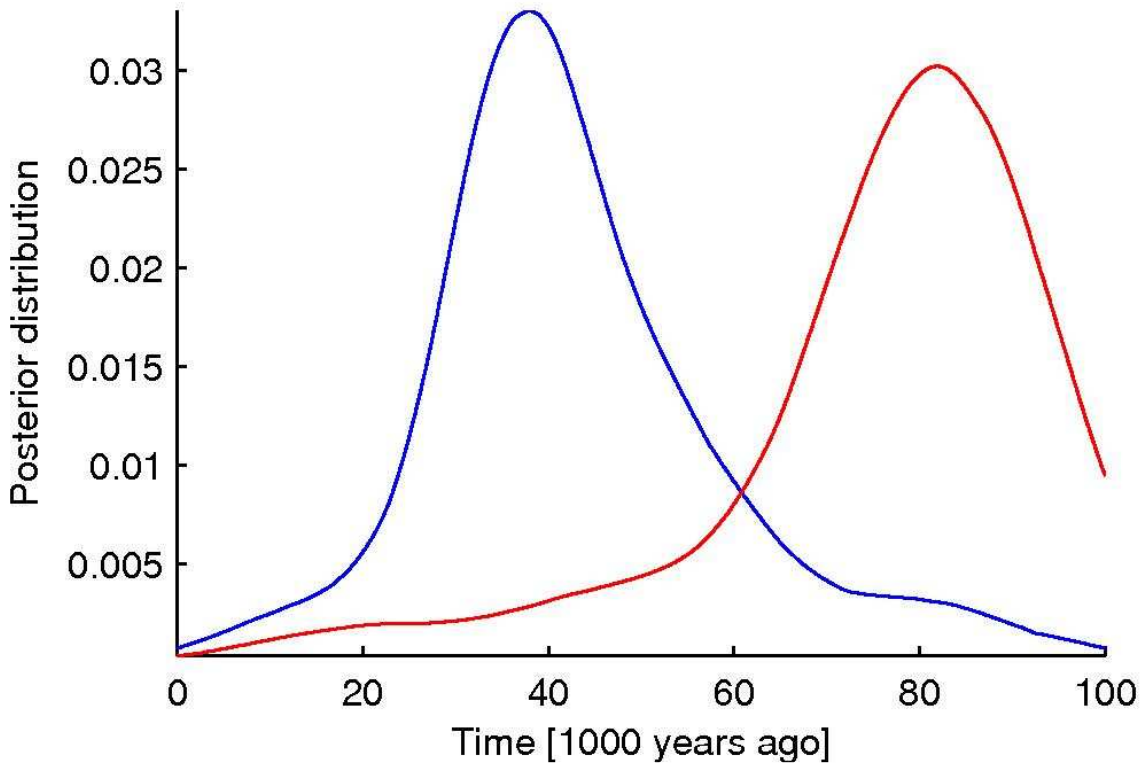

$r_M = 0.68$
$p < 0.001$
Pairwise genetic distance
0.00
0.05
0.10
0.15
0.20
0.25
0.30
3.5
4.0
4.5
5.0
5.5
log(geographic distance)

Figure 3


Posterior distribution
0.03
0.025
0.02
0.015
0.01
0.005
0
20
40
60
80
100
Time [1000 years ago]

Supplementary Table 1. Polymorphism of *P. falciparum* genes in the nine populations analyzed.

Locus		Africa		Southeast Asia		Oceania			South America		Worldwide
		Tanzania	Ghana	Thailand	Philippines	Papua New Guinea	Solomon Is.	Vanuatu	Brazil	Venezuela	
Pfcrt (1630 bp)	N	69	68	82	53	69	51	60	42	10	514
	S	30	13	10	4	8	4	3	9	3	49
	θ_π	0.00068 ± 0.00005	0.00042 ± 0.00006	0.00025 ± 0.00003	0.00024 ± 0.00003	0.00015 ± 0.00003	0.00010 ± 0.00004	0.00024 ± 0.00005	0.00073 ± 0.00006	0.00021 ± 0.00009	0.00041 ± 0.00002
	θ_S	0.00172 ± 0.00054	0.00085 ± 0.00033	0.00055 ± 0.00002	0.00024 ± 0.00013	0.00044 ± 0.00015	0.00024 ± 0.00014	0.00017 ± 0.00010	0.00058 ± 0.00025	0.00029 ± 0.00017	0.00196 ± 0.00029
	dS	0.00148 ± 0.00072	0.00103 ± 0.00071	0.00062 ± 0.00040	0.00049 ± 0.00036	0.00042 ± 0.00025	0.00030 ± 0.00023	0.00039 ± 0.00037	0.00111 ± 0.00072	0.00073 ± 0.00051	0.00088 ± 0.00042
	dN	0.00047 ± 0.00019	0.00026 ± 0.00011	0.00015 ± 0.00008	0.00018 ± 0.00018	0.00008 ± 0.00005	0.00005 ± 0.00004	0.00020 ± 0.00015	0.00063 ± 0.00028	0.00007 ± 0.00007	0.00029 ± 0.00009
		dS>dN: P= 0.090	dS>dN: P= 0.150	dS>dN: P= 0.133	dS>dN: P= 0.233	dS>dN: P= 0.095	dS>dN: P= 0.145	dS>dN: P= 0.332	dS>dN: P= 0.273	dS>dN: P= 0.102	dS>dN: P= 0.087
Pfdhfr (1413bp)	N	72	67	83	53	94	51	77	42	10	519
	S	7	1	2	2	3	2	1	1	1	14
	θ_π	0.00047 ± 0.00007	0.00059 ± 0.00008	0.00037 ± 0.00006	0.00038 ± 0.00003	0.00032 ± 0.00004	0.00050 ± 0.00006	0.00015 ± 0.00004	0.00036 ± 0.00002	0.00025 ± 0.00011	0.00040 ± 0.00002
	θ_S	0.00102 ± 0.00039	0.00119 ± 0.00045	0.00028 ± 0.00020	0.00031 ± 0.00022	0.00042 ± 0.00024	0.00031 ± 0.00022	0.00014 ± 0.00014	0.00016 ± 0.00016	0.00025 ± 0.00025	0.00145 ± 0.00039
	dS	0.00211 ± 0.00148	0.00269 ± 0.00179	0.00174 ± 0.00173	0.00185 ± 0.00171	0.00151 ± 0.00130	0.00245 ± 0.00181	0.00073 ± 0.00072	0.00176 ± 0.00175	0.00123 ± 0.00123	0.00190 ± 0.00160
	dN	0.00005 ± 0.00003	0.00048 ± 0.00047	0.00002 ± 0.00002	0	0.00002 ± 0.00002	0	0	0	0	0.00002 ± 0.00001
		dS>dN: P= 0.074	dS>dN: P= 0.061	dS>dN: P= 0.144	dS>dN: P= 0.128	dS>dN: P= 0.114	dS>dN: P= 0.083	dS>dN: P= 0.156	dS>dN: P= 0.145	dS>dN: P= 0.145	dS>dN: P= 0.080
Pfcrt-pfdhfr (6043 bp)	n	0.00063 ± 0.00004	0.00048 ± 0.00006	0.00029 ± 0.00003	0.00028 ± 0.00003	0.00020 ± 0.00003	0.00022 ± 0.00004	0.00022 ± 0.00004	0.00062 ± 0.00005	0.00022 ± 0.00008	0.00043 ± 0.00002
	θ	0.00153 ± 0.00025	0.00096 ± 0.00022	0.00048 ± 0.00014	0.00036 ± 0.00011	0.00043 ± 0.00013	0.00035 ± 0.00011	0.00016 ± 0.00008	0.00047 ± 0.00015	0.00028 ± 0.00014	0.00188 ± 0.00024

N: Number of sampled isolates
S: Number of segregating sites
θ_π: Average number of pairwise nucleotide differences
θ_S: Standardized number of polymorphic sites per site
dN: Mean number of synonymous substitutions
dS: Mean number of non-synonymous substitutions

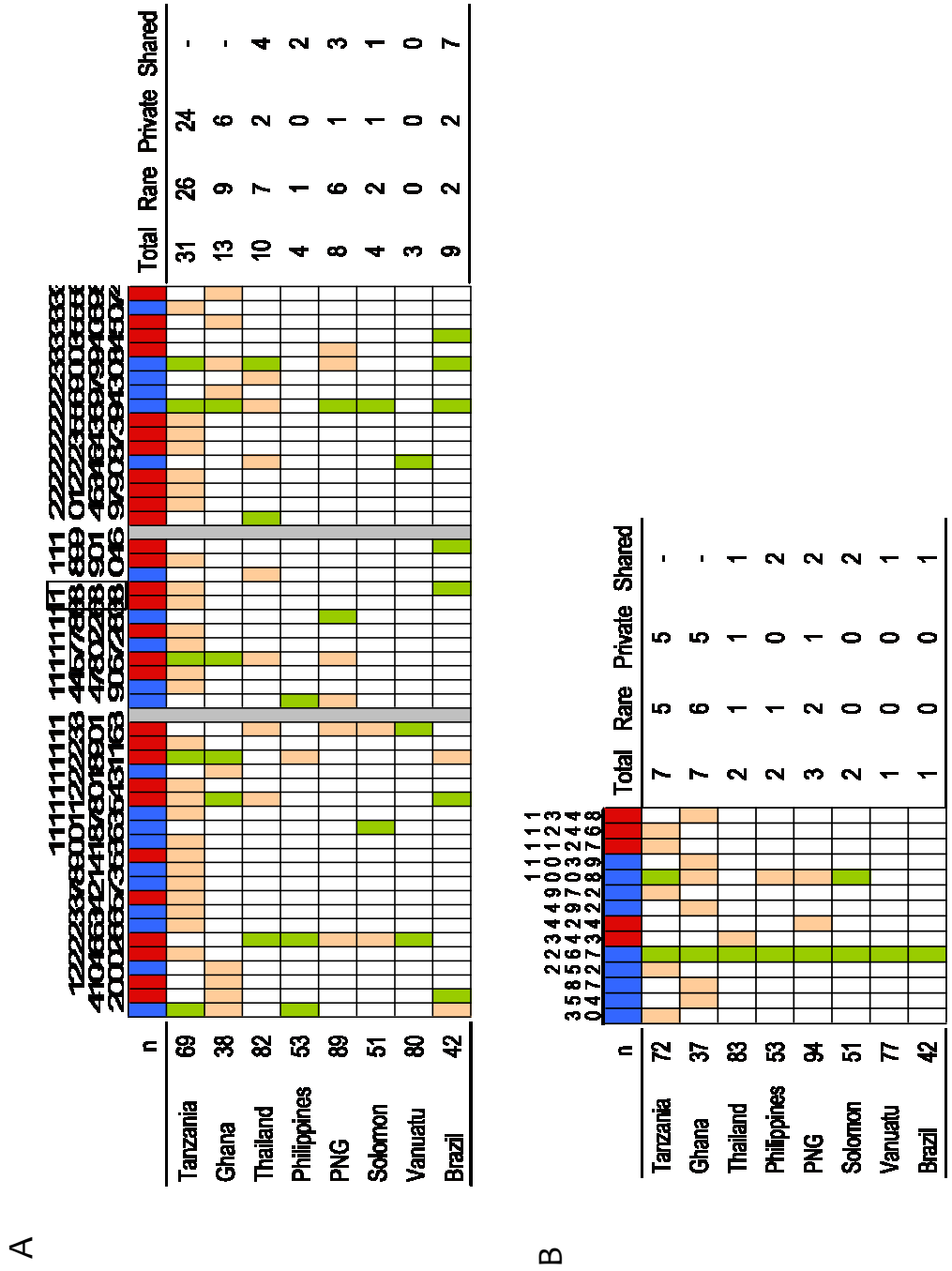
Supplementary Table 2. Entomological Infection rates (EIR) obtained from the literature

Country	Malaria endemicity	Entomological inoculation rate (bites/year)	Reference *
Tanzania	Hyper to holo	285	1, 2
Ghana	Hyper	418	1, 3
Thailand	Hypo	0.4 **	4
Philippines	Hypo to meso	5 **	5
Papua New Guinea	Meso to hyper	26 **	6
Solomon Islands	Meso to hyper	509	7
Vanuatu	Hypo to hyper	4 (in Santo island) **	8
Brazil	Hypo	Not available	
Venezuela	Hypo	Not available	

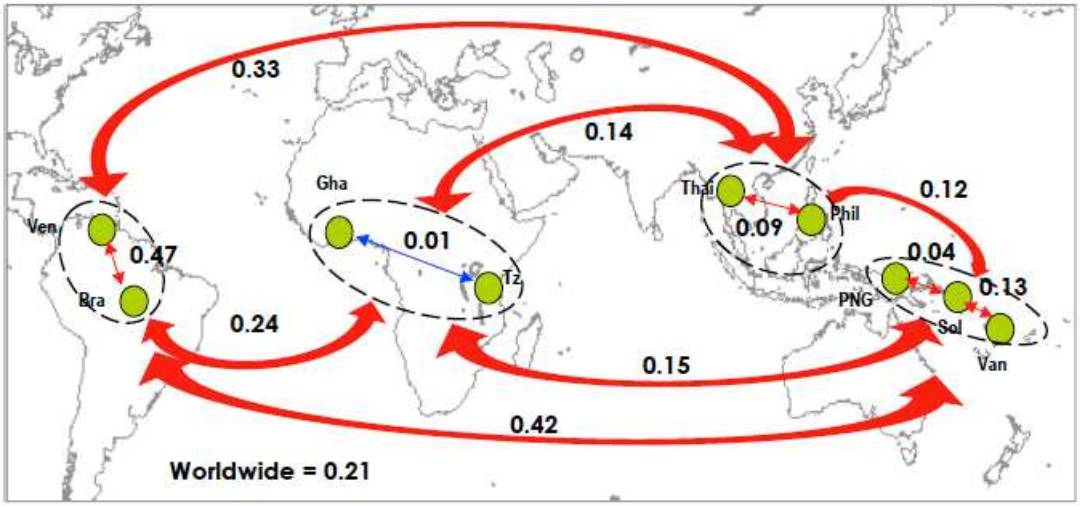
* References: 1, Hay SI, et al. (2000) Trans R Sco Trop Med Hyg 94:113-127; 2, Kelly-Hope LA and McKenzie FE Malaria J (2009) 8:19; 3, Gemperli, A. et al., (2006) Trop Med Int Health 11:1032-1046; 4, Lexemburger C, et al. (1996) Trans R Soc Trop Med Hyg 90:105-111; 5, Oberst RB, et al., (1988) Phil J Microbiol Infect Dis 17:41-45; 6, Benet A, et al. (2004) Am J Trop Med Hyg 71:277-84; 7, Hii JL. et al. (1993) Med Vet Entomol 7:333-338; 8, Williams RN, et al, (1995) Ann Trop Med Parasitol 89: 305-307. ** Values from the literature were halved in cases where numbers reported joint EIR for *P. falciparum* and *P. vivax*.

Supplementary table 3. Median values and 90% credibility intervals of the stepping stone parameters' posterior distributions, given a split between the *Plasmodium* species either at 2.5 or 6 Mya. The various parameters are defined in the methods.

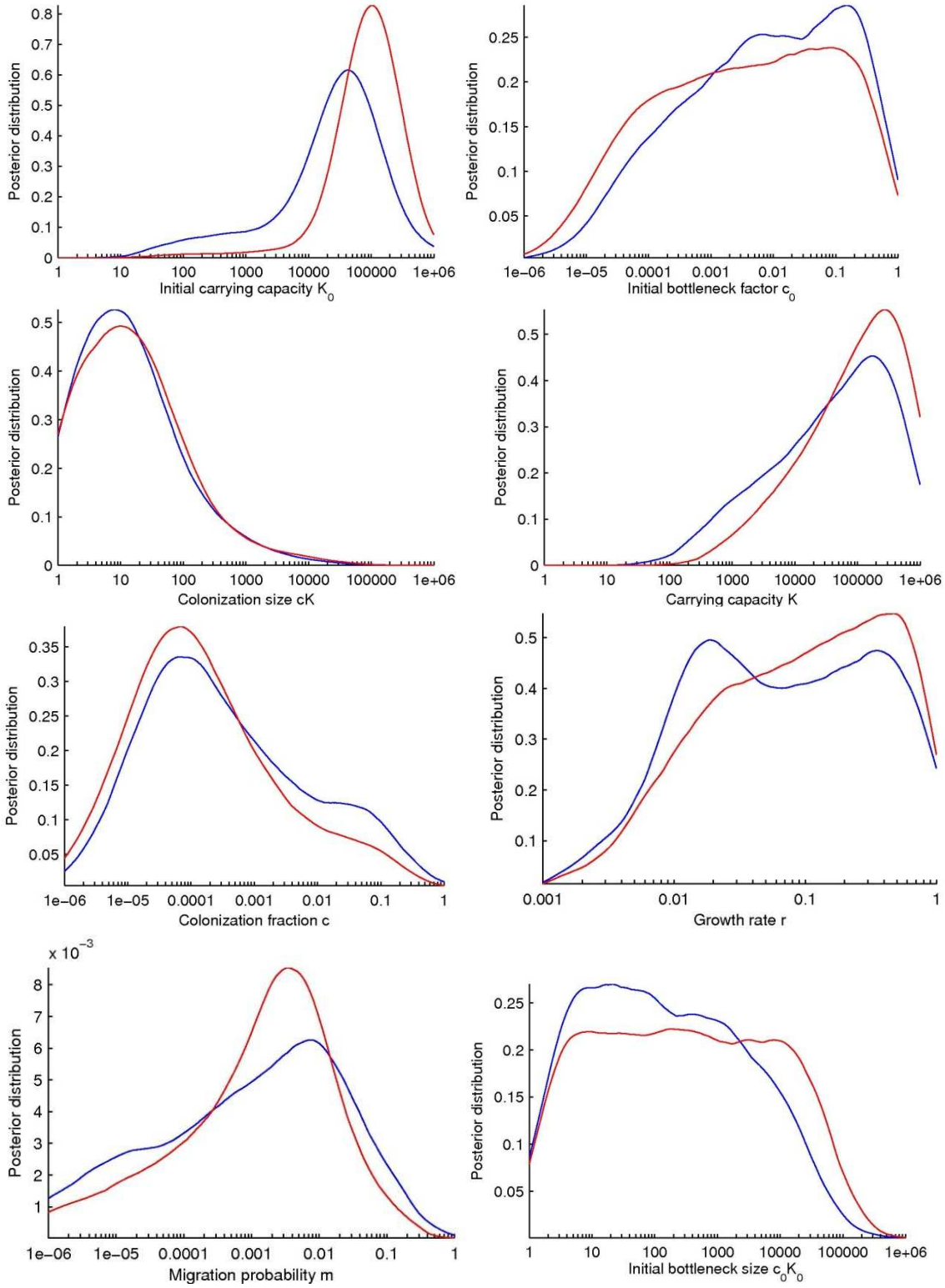
	Split 2.5 Mya			**Split 6 Mya**		
Parameter	**Median**	**90% CI**		**Median**	**90% CI**	
t	41,000	19,000	77,000	78,000	33,000	96,000
m	0.0016	4×10^{-6}	0.1	0.0020	6×10^{-5}	0.05
r	0.063	0.004	0.6	0.10	0.005	0.8
c	2.0×10^{-4}	5×10^{-6}	0.10	1.3×10^{-4}	4×10^{-6}	0.04
K	50,000	5×10^{2}	6×10^{5}	100,000	2×10^{3}	8×10^{5}
c_0	0.0080	3×10^{-5}	0.5	0.0040	2×10^{-5}	0.4
K_0	32,000	2×10^{2}	3×10^{5}	100,000	8×10^{3}	5×10^{5}
cK	13	1	200	16	1	250
c_0K_0	130	3	2×10^{4}	320	3	5×10^{4}



Supplementary Figure 1. Geographical distribution of SNPs in housekeeping genes of *P. falciparum*. SNPs in *serca* (A) and *adsl* (B) are shown for eight parasite populations from which >30 sequences were obtained. Polymorphic sites are shown in colored boxes: blue for synonymous and red for non-synonymous SNPs. SNPs with minor allele frequency (MAF) ≤5%, i.e., rare variants, are shown in green and those with MAF of >5% in beige. Monomorphic sites are non-colored. Positions of polymorphic sites are numbered after the 3D7 sequences (GenBank Accession numbers PFA0310c and PFB0295w). Private SNPs are those found only in one geographic area. Regions of tandem repeats are shown in grey boxes.



Supplementary Figure 2. Genetic differentiation between populations of *P. falciparum*. Red arrows indicate significant differentiation (F_{ST} significantly greater than zero; $p<0.05$) and blue arrows for non-significant differentiation.



Supplementary figure 3. Posterior distributions of the stepping stone parameters, given a split between the *Plasmodium* species at 2.5 (blue) and 6 (red) Mya.